75th MORSS CD Cover Page

712CD

For office use only 41205

UNCLASSIFIED DISCLOSURE FORM CD Presentation

12-14 June 2007, at US Naval Academy, Annapolis, MD

Please complete this form 712CD as your cover page to your electronic briefing submission to the MORSS CD. Do not fax to the MORS office.

Author Request (To be completed by applicant) - The following author(s) request authority to disclose the following presentation in the MORSS Final Report, for inclusion on the MORSS CD and/or posting on the MORS web site.

Name of Principal Author and all other author(s):

Joseph Parker, Ray Mitchell, Douglas Lowe - Teledyne Brown Engineering
Michael Galarneau, Peggy Fridlund, Martin Hill - Naval Health Research Center

Principal Author's Organization and address:

Teledyne Brown Engineering
300 Sparkman Dr. Huntsville, AL 35807

Phone: 256-726-1756

Fax: 256-726-2423

Email: joseph.parker@tbe.com

Original title on 712 A/B: Estimating Patient Condition Codes using Data Mining Techniques

Revised title:_____

Presented in (input and Bold one): (**WG_23**, CG___, Special Session ___, Poster, Demo, or Tutorial):

This presentation is believed to be:
UNCLASSIFIED AND APPROVED FOR PUBLIC RELEASE

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 JUN 2007	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Estimating Patient Condition Codes using Data Mining Techniques	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Teledyne Brown Engineering 300 Sparkman Dr. Huntsville, AL 35807**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM202526. Military Operations Research Society Symposium (75th) Held in Annapolis, Maryland on June 12-14, 2007, The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**19**	

TELEDYNE
BROWN ENGINEERING, INC.
A Teledyne Technologies Company

Estimating Patient Condition Codes using Data Mining Techniques

75th MORSS (WG 23)

Joseph Parker, Teledyne Brown Engineering
Ray Mitchell, Teledyne Brown Engineering
Douglas Lowe, Teledyne Brown Engineering
Michael Galarneau, Naval Health Research Center
Peggy Fridlund, Naval Health Research Center
Martin Hill, Naval Health Research Center

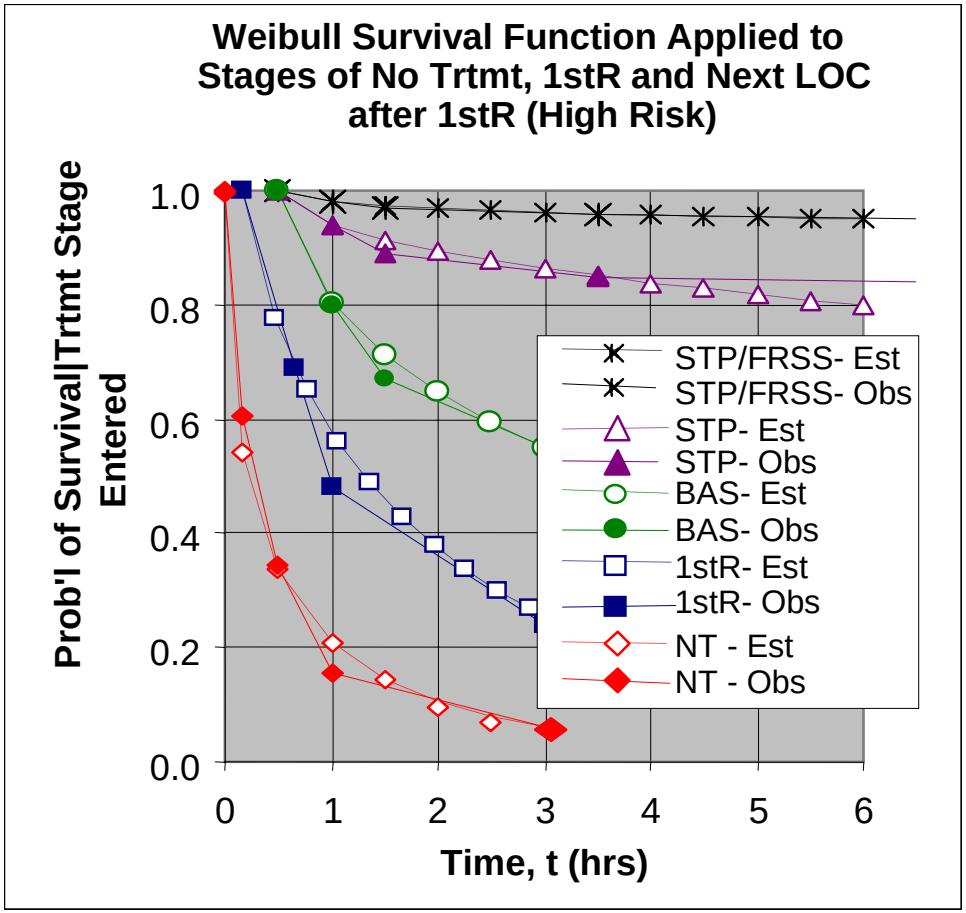
June 12, 2007

The Problem

TBE wanted to improve our TML+ mortality algorithm by factoring in the chance of death as a function of time based on real-world data.

The current DOW mortality algorithm is based on responses from a panel of medical SMEs rather than empirical data

Due to time limitations and the amount of data, we decided to pursue an automated method for assigning patient condition codes



Source Data

The Combat Trauma Registry (CTR) database maintained by the Naval Health Research Center records medical data from patients treated at Navy and Marine Corps medical facilities

Each CTR record documents a single visit (encounter) by a single patient

Each record contains over 50 individual fields of data, though often the data is incomplete

The most relevant data (such as injury descriptions) are primarily free text containing numerous abbreviations, medical terms, and misspellings

Example:
“FOREHEAD LAC X 2 NASAL FX DEFORMITY LT SUPERIOR ORBITAL RIDGE FX LT CORNEAL ABRASION WITH FOREIGN BODY~”

Analysis Method Benefits

Subject matter expert

Strengths

- True medical knowledge
- Recognition of misspelled words, abbreviations, and acronyms

Weaknesses

- Subjective – different SMEs may give different results
- Easy to miss important data
- Slow (several minutes per encounter)
- Expensive

Machine data mining

Strengths

- Inexpensive
- Consistent results
- Very fast

Weaknesses

- No real medical knowledge, only rules
- Accuracy is reduced by unusual words, abbreviations, acronyms
- Extremely difficult to develop algorithm

Step 1: Clean and Associate the Data

For privacy protection, NHRC replaced SSNs with another unique identifier

Each CTR record represents a single CTR form. One is filled out in the field after each patient encounter.

There may be multiple encounters (and forms) for a single person at different locations

A patient record was created to keep track of the relationship between multiple encounters and a single patient

fr_id	mtf	location	Pat_ID	mtf_receiv	Date of Injury
EF01423	FRSS2/STP4	AL QA'IM	P000001230	Camp Gannon	13-Apr-05
EF06508	SC Charlie	Camp FZ	P000001235	OP 3	24-Dec-05
EF01063	FRSS1/STP2	TA QUADDUM	P000001243	503 BAS	05-Jan-05
EF06485	SC Charlie	CAMP FZ	P000001290	FIELD	06-Jan-06
WF04531	BAS GAS 2 MHG	CAMP FALLUJAH	P000001293		04-Nov-05
EF00019	FRSS1/STP2	TQ	P000001378		
EF00573	BAS 2/1	FALLUJAH, IRAQ	P000001378	Mhg EFCAT	03-Apr-04
EF00054	FRSS1/STP2	Q	P000001398		
EF00395	FRSS1/STP2	OB-MAMUDIYAHA	P000001401	1ST LAR	08-Apr-04
EF01118	FRSS2/STP4	OB Al Qaim	P000001412		14-Apr-04

337	P000001373	12/23/2004 10:00:00 AM	15	37.5	23	33.33333333	
338	P000001378	4/3/2004 10:00:00 AM	15	67.5	16	55.125	

id	CTRFormID	MTF	TimeBegan	ConditionAt	ExtentOfTre	TimeTreatn	Deceased
405	EF00573	BAS 2/1	12:00:00 AM	Alert		12:00:00 AM	
406	EF00019	FRSS1/STP2	11:16:00 AM	Alert		12:30:00 PM	
0							

339	P000001382	11/9/2004 7:30:00 AM	53	50.2	54	50.2
340	P000001393	10/8/2005 3:00:00 PM	-1	0	-1	0

Step 2: Catalog Words

The algorithm begins by reading all text fields and cataloging every word and it's associated encounter record. For example, imagine this is in the injury description:

FOREHEAD LAC X 2 NASAL FX DEFORMITY LT SUPERIOR ORBITAL RIDGE FX LT CORNEAL ABRASION WITH FOREIGN BODY~

The algorithm generates the following initial word list:

- FOREHEAD
- LAC
- X
- 2
- NASAL
- FX
- DEFORMITY
- LT
- SUPERIOR
- ORBITAL
- RIDGE
- CORNEAL
- ABRASION
- WITH
- FOREIGN
- BODY

Step 3: Simplify/Consolidate Word List

The algorithm continues by eliminating common “noise” words and translating abbreviations it recognizes into full words. It also simplifies some words that are too specific for the PC code description. For instance, in this case, it changes “forehead” to “head” since there are no forehead-specific injuries in the PC list:

- FOREHEAD ⟹ HEAD
- LAC ⟹ LACERATION
- ~~X~~ (single-letter word)
- ~~2~~ (single-letter word)
- NASAL ⟹ NOSE
- FX ⟹ FRACTURE
- DEFORMITY
- LT ⟹ LEFT
- SUPERIOR
- ORBITAL ⟹ HEAD
- RIDGE
- CORNEAL ⟹ EYE
- ABRASION
- ~~WITH~~ (common word)
- FOREIGN
- BODY

Note one difficulty already:

- “X 2” was eliminated, even though it was specifying the number of lacerations, which may be important for diagnosis

FOREHEAD LAC X 2 NASAL FX DEFORMITY LT SUPERIOR ORBITAL RIDGE FX LT CORNEAL ABRASION WITH FOREIGN BODY~

Step 4: Associate Related Words

A specific list of adjectives are used by the algorithm to keep some important phrases together. In this case, “LEFT” is one of those words that must be kept with the word next to it, or it loses all relevance to the word matching algorithm. Phrases are treated the same as single words from this point on.

So, the word list becomes the following:

- HEAD
- LACERATION
- NOSE
- FRACTURE
- DEFORMITY
- LEFT SUPERIOR
- HEAD
- RIDGE
- LEFT EYE
- ABRASION
- FOREIGN
- BODY

Another difficulty:

- “LEFT” came before “SUPERIOR” so they were grouped together as a phrase, however the phrase “LEFT SUPERIOR” really describes the word “HEAD” after it. The algorithm currently does not support phrases longer than two words.

FOREHEAD LAC X 2 NASAL FX DEFORMITY LT SUPERIOR ORBITAL RIDGE FX LT CORNEAL ABRASION WITH FOREIGN BODY~

Step 5: Assign Weights

Each word (or phrase) is assigned a numeric value based on the location in which it was found. For instance, words from the injury description have a higher weight than words from the SOAP notes, which tend to tell more about the treatment than the injury. Multiple instances of the same word are counted separately then added together.

Also some words are given higher weights because they are highly relevant when determining the patient condition (these are usually related to the anatomical location).

Examples:

Word	Process	Weight
HEAD	(0 + 100 (found in inj desc) + 100 (highly relevant word)) X 2 (two instances)	400
RIDGE	(0 + 100 (found in inj desc))	100

FOREHEAD LAC X 2 NASAL FX DEFORMITY LT SUPERIOR ORBITAL RIDGE FX LT CORNEAL ABRASION WITH FOREIGN BODY~

Step 6: Assign Category

- PC descriptions are parsed in a similar manner
- "Instant category match" words are considered (such as HEAD). These limit the PC choice to those in the head category
- Anatomical location is considered to further limit PC category, if possible
- Keywords (such as KIA, CPR, Intubated, etc.) used to assign definitely/probably/maybe life threatening to the patient.
- Category is further limited to only PCs that are LT if the patient is "definitely" LT.

Anatomical record fields checked "yes"	Resulting PC Category
* Neck, Head, Face, Eye, or Ear	Head
* Genitalia, Abdomen, or Pelvis	Abdomen & Pelvis
* Back	Spine
* Thorax/Back or Chest	Thorax
Lower Extremity	Lower Limbs
Upper Extremity	Upper Limbs
Two or more of those criteria marked with * above	Multiple Injury Wounds

FOREHEAD LAC X 2 NASAL FX DEFORMITY LT SUPERIOR ORBITAL RIDGE FX LT CORNEAL ABRASION WITH FOREIGN BODY~

Word	Weight
HEAD	400
LACERATION	100
NOSE	200
FRACTURE	200
DEFORMITY	100
LEFT SUPERIOR	100
RIDGE	100
LEFT EYE	200
ABRASION	100
FOREIGN	100
BODY	100

Word/weight list for our sample

Step 7: Generate PC Match Ranking List

Now, each encounter's word list is compared to each PC description word list. The weights for all matches are added together and divided by the number of relevant words in the PC description. This keeps long PC descriptions from matching more often than short ones simply because they have more words.

The top three match indexes are reported to the analyst.

PC Description Word List (PC 10)

Word	Weight
HEAD	300
CONTUSION	100
OPEN	200
FRACTURE	200
MODERATE	100
NO HEAD	100
FRAGMENT	100
DEPRESSED	200

Example Encounter Word List

Word	Weight
HEAD	400
LACERATION	100
NOSE	200
FRACTURE	200
DEFORMITY	100
LEFT SUPERIOR	100
RIDGE	100
LEFT EYE	200
ABRASION	100
FOREIGN	100
BODY	100

300+400+200+200 = 1100
1100 / 8 = 137.5

Our encounter matches PC 10 with a confidence index of 137.5.

FOREHEAD LAC X 2 NASAL FX DEFORMITY LT SUPERIOR ORBITAL RIDGE FX LT CORNEAL ABRASION WITH FOREIGN BODY~

How did it do?

The algorithm was designed against a list of 53 CTR patient records that had already had a PC assigned by NHRC.

Data Set	Category Match %	Top 3 PC Match %
53 NHRC records (already had PC assigned by NHRC)	91%	45%
37 records algorithm determined to be abdomen & pelvis category and “defintely” LT	89%	70%
141 MIW records	~69%*	41%

* Counted all definite matches, “right category/wrong PC” matches, and “right category, NLT” matches

During development, we determined there was insufficient time to get exact PCs and started trying just for category matches.

MIW was the most difficult category for the estimator to compare, since the anatomical location is less relevant in choosing the proper PC.

Estimator picked the correct PC as one of it’s top three almost 50% of the time. Not adequate for our analysis, but clearly has some value as a first step in an automated methodology.

Conclusion - Potential Improvements

- Decrease reliance on individual fields (like anatomical location) in favor of free-text fields like injury description. They are harder to parse but relying too much on individual fields may be misleading if data is bad.

- Add adjective recognition, keep adjectives with the words they describe. Support complex descriptive sentence structures, like “superficial 3-inch thigh laceration.”

- Severity ranking – attempt to determine severity of injuries and match with only the most severe

- Add “medical visualization” allowing algorithm to understand that some types of injuries, due to their location, may impact nearby organs.

- Add a rudimentary expert system with some medical knowledge rules. For instance, if blood was used, the injury was probably life threatening.

- Expand to more complex coding systems, like ICD-9 and ICD-10. This could be used for medical coding or for building a PC to ICD mapping.

Conclusion – Contact information

Email: joseph.parker@tbe.com
For a more detailed technical description of the algorithm, word lists, etc.

Visit www.tmlsim.com
For more information on TML+ and the DOW algorithm.

Backups

Extracting Mortality Related Data from the CTR File

Receive CTR data from NHRC, extract patients and encounters, fix date formatting (3790 encounters)

Keep only records with adequate timing data (2961 encounters)

Perform automated LT analysis, keep only LT records (554 encounters)

Give patient/encounter list to SMEs, asking for NLT, High, Med, Low, Not enough info rating for each patient (436 patients, 554 encounters)

Reconcile and merge SME results, export Excel worksheet for analysis (Resulting patients: 158 H, 42 M, 44 L, 192 Other)

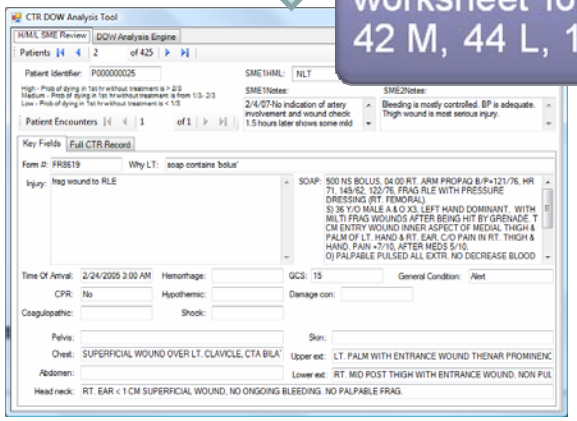
Adequate Timing Data?			Kept for DOW study
Injury	Arrival	Disposition	
No	No	No	No
No	Yes	No	No
No	No	Yes	No
No	Yes	Yes	Yes
Yes	No	No	No
Yes	Yes	No	Yes
Yes	No	Yes	Yes
Yes	Yes	Yes	Yes

Injury time, arrival time, and disposition time are available in the CTR data, but not every encounter has all data. Records were kept based on the criteria in the table above.

Automated LT algorithm determined likely LT encounters based on criteria from 12/12/06 NHRC and TBE technical interchange meeting.

Criteria consisted of words and phrases found anywhere in the encounter text (such as “Pos FAST” or “Deceased”) and qualifiers in particular fields (such as Hemorrhage = II, III, or IV). See full algorithm description document for more details.

SMEs were provided an application to evaluate LT status of each patient



Algorithm is described in detail, including all LT qualifiers, in the 4/11/2007 document “DOW PC Estimator Algorithm Summary 4-11-07.doc”

Patient Condition Code Sample

PC	Description
10	CEREBRAL CONTUSION WITH OPEN SKULL FRACTURE MODERATE - WITHOUT INTRACRANIAL FRAGMENTS AND/OR DEPRESSED SKULL FRACTURE
11	INTRACRANIAL HEMORRHAGE SPONTANEOUS NONTRAUMATIC ALL CASES
13	WOUND SCALP OPEN WITHOUT CEREBRAL INJURY OR SKULL FRACTURE SEVERE - SCALPED WITH AVULSION OF TISSUE
14	WOUND SCALP OPEN WITHOUT CEREBRAL INJURY OR SKULL FRACTURE MODERATE - SCALP LACERATION
15	FRACTURE FACIAL BONES CLOSED EXCLUSIVE OF MANDIBLE SEVERE - MULTIPLE FRACTURES
16	FRACTURE FACIAL BONES CLOSED EXCLUSIVE OF MANDIBLE MODERATE - SINGLE FRACTURE
17	WOUND FACE JAWS AND NECK OPEN LACERATED WITH ASSOCIATED FRACTURES EXCLUDING SPINAL FRACTURES SEVERE - WITH AIRWAY OBSTRUCTION
18	WOUND FACE JAWS AND NECK OPEN LACERATED WITH ASSOCIATED FRACTURES EXCLUDING SPINAL FRACTURES MODERATE - WITHOUT AIRWAY OBSTRUCTION; EYELID AND EYEBALL LACERATION WITH RETAINED INTRAOCULAR FOREIGN BODY
19	WOUND FACE AND NECK OPEN LACERATED CONTUSED WITHOUT FRACTURES SEVERE - WITH AIRWAY OBSTRUCTIONS AND/OR MAJOR VESSEL INVOLVEMENT
20	WOUND FACE AND NECK OPEN LACERATED CONTUSED WITHOUT FRACTURES MODERATE - WITHOUT AIRWAY OBSTRUCTION OR MAJOR VESSEL INVOLVEMENT
21	EYE WOUND SEVERE - LOSS OF INTRAOCULAR FLUID WITH/WITHOUT RETINAL DETACHMENT, WITH SEVERE LID LACERATION, EYE NOT SALVAGEABLE
22	EYE WOUND LACERATED MODERATE - WITHOUT RETINAL DETACHMENT OR RETINAL INJURY NO FOREIGN BODY RETAINED WITHOUT LOSS OF VITREOUS FLUID PATIENT HAS HYPHEMA EYE SALVAGEABLE
23	HEARING IMPAIRMENT SEVERE
24	HEARING IMPAIRMENT MODERATE
25	FRACTURE SPINE CLOSED WITHOUT CORD DAMAGE UNSTABLE LESION
26	FRACTURE SPINE CLOSED WITHOUT CORD DAMAGE STABLE LESION
27	FRACTURE SPINE CLOSED WITH CORD DAMAGE CERVICAL SPINE WITH RESPIRATORY INVOLVEMENT
28	FRACTURE SPINE CLOSED WITH CORD DAMAGE BELOW CERVICAL SPINE (PROGRESSIVE)
29	FRACTURE SPINE OPEN WITH CORD DAMAGE CERVICAL SPINE WITH RESPIRATORY DISTRESS

Patient Condition Categories

- Abdomen & Pelvis
- Battle Fatigue
- Burns
- Cardiovascular
- Directed Energy Weapon Eye Lesion
- Environmental
- Eye/Ear Disease
- Female Specific
- Gastrointestinal
- General
- Genitourinary
- Head
- Infectious/Parasitic
- Lower Limbs
- Dental
- Multiple Injury Wounds
- Miscellaneous
- Neuropsychiatric
- Not Assigned
- Preventive Medicine
- Respiratory
- Sexually Transmitted Disease
- Spine
- Sprains & Strains
- Superficial/Soft Tissue
- Surgical
- Thorax
- Upper Limbs
- Dermatological
- Nuclear, Biological, Chemical